Can Twitter Analytics Predict Election Outcomes? An Insight from 2017 Punjab Assembly Elections

Prabhsimran Singh,
Department of Computer Engineering & Technology,
Guru Nanak Dev University, Amritsar, India.
prabh_singh32@yahoo.com

Yogesh K. Dwivedi,
School of Management, Emerging Market Research Center (EMaRC),
Swansea University, Swansea, UK.
y.k.dwivedi@swansea.ac.uk

Karanjeet Singh Kahlon,
Department of Computer Science,
Guru Nanak Dev University, Amritsar, India.
karankahlon@gndu.ac.in

Annie Pathania,
Department of Computer Engineering & Technology,
Guru Nanak Dev University, Amritsar, India.
apathania89@gmail.com

Ravinder Singh Sawhney,
Department of Electronics Technology,
Guru Nanak Dev University, Amritsar, India.
sawhney.ece@gndu.ac.in

Abstract: Since the beginning of this decade, there has seen an exponential growth in number of internet users using social media, especially Twitter for sharing their views on various topics of common interest like sports, products, politics etc. Due to the active participation of large number of people on Twitter, huge amount of data (i.e. big data) is being generated, which can be put to use (after refining) to analyze real world problems. This paper takes into consideration the Twitter data related to the 2017 Punjab (a state of India) assembly elections and applies different social media analytic techniques on collected tweets to extract and unearth hidden but useful information. In addition to this, we have employed machine learning algorithm to perform polarity analysis and have proposed a new seat forecasting method to accurately predict the number of seats that a political party is likely to win in the elections. Our results confirmed that Indian National Congress was likely to emerge winner and that in fact was the outcome, when results got declared.

1. Introduction

Social Media consist of different platforms (for example Facebook, Instagram, LinkedIn, Twitter, WhatsApp, Youtube etc) which provide digital environment, where people from different backgrounds and regions communicate on common interests (viz: Business, Economic, Political & Social Issues) (Alalwan et al., 2017; Aladwani and Dwivedi, 2018; Bertot et al., 2012; Gintova, 2019; Dwivedi et al., 2015; Kamboj et al., 2018; Kapoor et al., 2018; Khan and Khan, 2019; Rad et al., 2017; Tang et al., 2019; Zhang et al., 2019). Talking about politics, any political event attracts a lot of social media attention. When it comes to the elections, which is an important event for any country, region or state, the rate of social media buzz created by such events is very high and often attracts a lot of people. Such popularity suggests that social media's role is an important political game changer (Howard and Parks, 2012; Kavanaugh et. al., 2012; Lim, 2012; Yaqub et. al., 2017).

Most of the researchers have described social media as a mode for large scale communication (Alalwan et al., 2017; Boynton and Richardson, 2016) and have also outlined its role in impacting the outcome of election events (Bruns and Stieglitz, 2013; Grover et al., 2019). Barack Obama's 2008 US presidential elections (Bruns and Stieglitz, 2013) and Narendra Modi's 2014 Indian General Elections (Kapoor and Dwivedi, 2015) are two of the biggest case studies which clearly portrays how social media turned out to be the biggest driving factor in their respective campaigns towards victories (Grover et al., 2019). Voters definitely get empowered by social media blitz which enhances deliberative democracy among them (Lawrence et al., 2010; Yardi and Boyd, 2010). Deliberations may eventually help the voters in: (a) refining their own opinions; (b) listening to different opinions; and (c) identifying common ends and means (Lawrence et al., 2010).

Ever since the 2014 Indian general elections, there has been a significant rise in the use of social media (especially Twitter) in India by politicians as well as common people (voters) (Kapoor and Dwivedi, 2015). Every subsequent election in India has seen a significant use of Twitter by the politicians for campaigning, while the voters use social media to show their support as well as hindrance towards a particular candidate or political party. Similar trend and observations have been made in the case of 2017 Punjab (India) assembly elections, which was bound to witness a triangular battle for the first time with the addition of a new political outfit (namely, Aam Aadmi Party (AAP)) contesting against the traditional rivals Indian National Congress (INC) and the alliance of Shiromani Akali Dal (SAD) & Bharatiya Janata Party (BJP).

Drawing motivation from above, the aim of this paper *is to explore the role of Twitter in 2017 Punjab assembly elections for understanding how Twitter can serve as an important tool to predict the outcome of future political events when combined with techniques for social media analytics (SMA) techniques such as sentiment analysis, content analysis, topic modeling* etc. In addition to this, we have attempted to evaluate the possible change in voting preference as the elections drew near. Further, we have proposed a new seat forecasting technique using machine

learning algorithm and mathematic notations that can predict (based on Twitter conversations about elections) the possible number of seats a political party is likely to win in the elections. In order to accomplish the proposed objectives, we have collected a total of 9,157 tweets over a period of 28 days and applied social media analytics to analyze the collected data.

The remaining part of this paper is structured as follows: Section 2 gives a brief background of Punjab assembly elections. Section 3 discuses related work done in this field. This followed by various pre-poll factors affecting the election results in Section 4. In Section 5, research methodology has been discussed briefly. Section 6 gives the details of data collection, while Section 7 deals with the data preprocessing phase. Section 8 discusses the various social media analytics techniques along with its results followed by a seat forecasting method in Section 9. A discussion of results is presented in Section 10. Finally, we conclude in Section 11 by outlining key points and limitations of this study.

2. Background Punjab Assembly Elections

Ever since the formation of modern day Punjab (India) in 1966, elections in Punjab had always been a two horse race. Two traditional parties INC and SAD along with their alliance partners BJP had always ruled the state for five years on alternate terms. This trend was broken in 2012 Punjab assembly elections when the ruling SAD were able to halt the trend to retain the power and eventually becoming the first party since the creation of modern Punjab to rule the state for two successive terms (ECI, 2018a).

During 2014 Indian general elections, the people of Punjab had an opportunity to vote for a newer political entity AAP. As a result, AAP won 4 out of the 13 Lok Sabha (Parliamentary) seats in Punjab while it drew nil in rest of India (ECI, 2014). Since then, there have been speculations that during the 2017 Punjab Assembly Elections, AAP would be a force to reckon along with traditional parties thus making the contest triangular. Going into 2017 Punjab Assembly Elections, the ruling SAD-BJP government was facing huge anti-incumbency which got triggered during 2014 Indian General Elections and saw even the current Finance Minister Arun Jaitley becoming the victim by failing to win in Amritsar Parliamentary Constituency considered, which traditionally have been a safe heaven for the combine (ECI, 2018b). The outcome was nothing less than the tsunami for the SAD-BJP government that had ruled the state since 2007 and Amritsar was considered a strong bastion of the SAD-BJP combine.

The election commission of India (ECI) is solely responsible for conducting the state legislative elections in India (ECI, 2018a). Currently the Punjab legislative assembly has 117 members, which are elected for the tenure of 5 years. Out of the 117, 59 seats are required by a political party in order to establish a majority government. The ECI conducted assembly polls in Punjab on February 4, 2017 and results were declared on March 11, 2017.

3. Literature Review

This section is divided into four sub sections, namely social media, use of social media by politicians as well as politics, election prediction using Twitter and finally the presentation of the gaps, research questions and contributions made to the study.

3.1. Social Media

In this current decade, the popularity of social media especially Twitter has increased dramatically on the global stage (Panagiotopoulos et al., 2016; Statista, 2018). This exponential increase in usability has provided the researchers with a large amount of user generated data in order to predict or track public opinions (Blazquez and Domenech, 2017; Mellon, 2014; Ortiz et al., 2011; Preis et al., 2013; Suh, 2015). There has been an exponential rise in the traffic related to Twitter data (tweets) for predicting the outcome of future events, but work done is still in its infancy and encounters problems in one way or the other way.

Twitter has been extensively used for multiple purposes namely promoting various public policies (Harris et al., 2014), warning for natural disasters (Chatfield et al., 2013), raising environment issues (Cody et al., 2015), evaluating voting intentions (Ceron et al., 2014) and handling emergency situations (Panagiotopoulos et al., 2016). The huge potential of Twitter to uncover unbiased information from user generated data was the foremost factor for using Twitter in our study.

3.2. Social Media and Politics

Many politicians across the nations use social media platforms (especially Twitter) for sharing their communiqué among their followers (Kelm et al., 2017) and being a multi-directional approach, it offers an effective as well as interactive communication (Kruikemeier et al., 2016). The political information shared on social media mostly engages young people (Vromen et al., 2015). The published literature indicates that the online engagement on social media impacts user's sentiments (Ibrahim et al., 2017) which further influences the outcome of general elections in emerging economies such as India (Safiullah et al., 2017).

Twitter has been used sparingly by politicians for advertising (Domingo and Martos, 2015; Hutchins, 2016), broadcasting purposes (Hutchins, 2016; Theocharis et al., 2016) and for engaging with citizens (Ahmed et al., 2016). LaMarre and Suzuki-Lambrecht (2013) have been able to show that Twitter usage by the politicians increases their chances of winning an election. The more the politician is active on social media, the more journalists as well as press give publicity to the news aired by the politician (Rauchfleisch and Metag, 2016).

3.3. Election Prediction using Twitter

Tumasjan et al. (2010) were the first to use Twitter to estimate the results of 2009 German federal elections. The major limitation of this study was that it was carried out post-election.

Though the authors claimed high accuracy rate for top six political parties by using the simple technique of counting the number of tweets for each party, this technique attracted huge criticism (Gayo Avello et al., 2011; Gayo Avello, 2012; Jungherr et al., 2012). Gayo Avello (2011, 2012) pointed out the requisition to make a true prediction and that was the prediction must be made before the actual event has taken place. Moreover, he emphasized that simply relying on the number of tweets is not an accurate way of making the prediction and some computational model should be included in order to make the election predictions.

Subsequent studies (DiGrazia et al., 2013; Jungherr et al., 2013; Makazhanov et al., 2014; Sang et al., 2012) followed the advice of Gayo-Avello and added some computational approach to their work rather than simply relying on the number of tweet mentions towards making the predictions. Ceron et al. (2014) used sentiment analysis along with the tweets to check the popularity of political candidates in Italian parliamentary elections 2011 and French presidential Elections 2012, while Singh et al. (2017) used sentiment analysis & tweets to forecast the outcome of 2016 US Presidential elections. Burnap et al. (2016) & Srivastava et al. (2015) improvised this research approach by introducing the mechanism to predict the number of seats, the contesting parties are likely to win in the elections using sentiment analysis of the tweets collected from Twitter. The prediction made by Srivastava et al. (2015) proved to be correct for 2015 Delhi assembly elections, while Burnap et al. (2016) made an incorrect prediction for 2015 UK General Elections. The above literature establishes a fact that Twitter can be utilized as a powerful tool for making electoral predictions if a proper methodology based on certain mathematical tools is also included rather than simply relying on the number of tweet mentions.

3.4. Gaps and Contributions

Till date to the best of our knowledge no study has presented a generic technique that can predict the outcome of an election with precise number of seats. Further, since Twitter is a multi-lingual platform, many people post their opinions in their native language (i.e. Non English language), but almost negligible work has been done in order to take into consideration the opinions posted in their native language. So through this study, we have tried to answer the following questions:

(a) Can a generic technique be developed using Twitter data that can predict the election outcome with precise number of seats?
(b) How non English tweets can be evaluated to predict the outcome of an event?

Our research work is motivated by the work & guidance as suggested by Burnap et al. (2016); Gayo Avello (2011, 2012) and Srivastava et al. (2015). We first applied have social media analytics including sentiment analysis to tweets collected from Twitter in order to get their polarity and calculated the sentiment scores. Then we have applied proposed seat forecasting method in order to make the prediction for the number of seats that a party was likely to win in 2017 Punjab assembly election. The combination of these two approaches helped us to predict the event outcome in close proximity to the actual results.

4. Pre-Poll Factors

As the 2017 Punjab assembly elections approached, in addition to the personal voting preferences there were several factors that were affecting the voting choice of the voters in one way or the other. These factors are discussed as following:

a) **BJP Winning streak:** Ever since the landslide victory of BJP in 2014 Indian General Elections (ECI, 2014), BJP had an incredible winning streak in various state assembly elections (Andhra Pradesh, Arunachal Pradesh, Pondicherry, Sikkim, Haryana, Maharashtra, Jammu and Kashmir, Jharkhand and Assam) to form government by its own or with its alliance partners. So coming into 2017 Punjab Assembly Elections, BJP and its alliance partner SAD were in seventh heaven, having ruled the state for last one decade.

b) **Anti-incumbency against SAD:** Heading into 2017 Punjab Assembly Elections, the ruling party SAD was facing huge anti-incumbency (Times of India Blog, 2016a). This anti-incumbency was a result of various issues like unemployment, farmer loans & suicides, increasing crime rate and drug peddling in the state.

c) **Rise of AAP in Punjab:** AAP made a dream debut in politics forming a government in New Delhi, as a result of 2013 Delhi Legislative Assembly Elections (Times of India, 2013). However, its ambition to become a national party suffered a huge setback as a result of 2014 Indian General Assembly as they won only 4 out of 543 seats (ECI, 2014). However, all these 4 seats were won from Punjab state. This result made clear that AAP is providing people of Punjab with an alternative heading into 2017 Punjab Assembly Elections to traditional parties like SAD-BJP and INC. However, AAP was constantly in the news for various infightings among its leaders (Business Standard, 2016). To make the matter worse, AAP was unable to form an alliance with Navjot Singh Sidhu, which many newspapers claimed to be an invincible combination (Times of India Blog, 2016b). Ahead of the elections, AAP formed an alliance with Simarjit Singh Bains led Lok Insaaf Party (LIP) (Punjabi Tribune, 2016), a popular political leader from Ludhiana district. Under the alliance, it was decided that AAP will be contesting on 112 seats while be LIP will contesting on the remaining 5 seats.

d) **INC's Aggressive Approach**: Unlike BJP, who were having an incredible winning streak, INC lost many states they were ruling. So INC came into these elections with aggressive approach knowing that they have nothing to lose (Firstpost, 2016a). INC first made Prasant Kishore (Election Strategist) in charge of election campaign (ABP News, 2016). Kishore introduced initiatives like "Coffee with Captain", "Halke Vich Captain" and job scheme ("Har Ghar Ton Ikk Captain") (Hargharcaptain Website, 2018; Times of India, 2016) to make sure that the Chief Minister candidate Captain Amrinder Singh connects to the urban and rural population of Punjab in order to enhance its popularity (Indian Express, 2016). Seeing this approach, many leaders from other parties (AAP, BJP & SAD) (News 18, 2016)

including Cricketer turned politician Navjot Singh Sidhu, joined INC party (Hindustan Times, 2017; Times of India, 2017).

e) **Sidhu Factor**: Cricketer turned politician Navjot Singh Sidhu is a role model for youngsters and enjoys a huge fan following due to his clean political image. Sidhu was always a BJP loyal, but rift between the alliance partners SAD grew because of the various political issues. Ever since Sidhu was denied the Lokh Sabha ticket from Amritsar constituency from where he was sitting MP (Member of Parliament) for the last 3 terms. There were speculations that Sidhu was not happy as BJP continues its alliance with SAD. Sidhu was given Rajya Sabha MP seat in April, 2016 in order to end this growing rift (Tribune, 2016). However, within 2 months he resigned from Rajya Sabha and finally in September, 2016 he quit the primary membership of BJP. During all this time, there were rumors that Sidhu might join AAP as their Chief Minister (CM) Candidate and many newspapers and political pandits termed it as an invincible combination (Firstpost, 2016b). Ultimately in January, 2017, Sidhu joined INC ending all the speculations (Hindustan Times, 2017; Times of India, 2017).

5. Research Methodology

The aim of this paper is to do in depth analysis of 2017 Punjab assembly elections using social media analytics. So in order to accomplish this mammoth task, we follow the methodology as given in Algorithm 1.

Algorithm 1

Start

1: Collecting the tweets from Twitter based on specific hashtags (#) for each political party.

2: Preprocessing each tweet.

2.1: Removing the unrelated/ unwanted tweets.

2.2: Converting Punjabi language tweets into English language

2.3: Converting all the tweets into lower case.

2.4: Removing all punctuations.

2.5: Removing all the numbers (numeric data).

2.6: Removing all the English stop words.

2.7: Removing all the web links & HTML tags.

2.8: Removing all the extra blank spaces.

3: Performing Social Media Analytics

3.1: Calculating tweet statistics.

3.2: Performing tweet network analysis.

3.3: Performing # hashtag & @ mentions analysis.

3.4: Performing sentiment analysis.

3.4.1: Performing e-motion analysis.

3.4.2: Performing polarity analysis.

4: Based on polarity analysis, calculating the seat share (i.e. number of seats) using the proposed method.

End

We have divided our procedural methodology into three phases. Phase-1 starts with the collection of tweets from Twitter to form a dataset. Tweets are collected in a time specific manner based upon specific hashtag (#). Once the collection of the tweets is complete, phase-2 starts where we first check the presence of any non-english tweets. If any non-english tweet is present in the collected dataset, then we convert that tweet into English language in such a way that its semantic meaning is preserved. Once all the tweets in the dataset are in English language we move ahead to the task of data preprocessing. The task of data preprocessing is done in order to avoid any ambiguity that may alter the desired results.

Once our coordinated dataset is ready, we have applied various social media analytics techniques (Chang, 2017; Singh et al., 2019; Stieglitz and Dang-Xuan, 2013). Generally, social media analytics is categorized into four broad categories namely descriptive analysis, content analysis, network analysis and geospatial analysis (Singh et. al., 2018a). The descriptive analysis gives descriptive statistics about number of tweets, number of tweet senders, hashtags, @mentions, word cloud etc (Chae, 2015; Ganis and Kohirkar, 2015; Purohit et al., 2013; Singh et al., 2018b). The content analysis deals with obtaining semantic content from the given text. This is achieved using techniques like sentiment analysis, topic modeling etc (Kassarjian, 1977; Kayser and Blind, 2017; Llewellyn et al., 2015; Zhang et al., 2016). The network analysis allows us to identify different communities and cluster of users based upon their opinion towards an entity (Abascal-Mena et. al., 2015; HerdaĞdelen et al., 2013; Stieglitz and Dang-Xuan, 2013). Finally, a geospatial analysis deals with location based and time series analysis (Attu and Terras, 2017; Saboo et al., 2016; Singh et al., 2018c; Stephens and Poorthuis, 2015). All these techniques produce differentiated results that together provided us with useful insights from different perspectives.

We have used supervised machine learning algorithms to perform the task of polarity analysis. Finally, in phase-3 based on the polarity of each tweet, we have calculated seat share (i.e. number of seats) for all the three main political parties AAP-LIP, SAD-BJP and INC as other political parties virtually stands no ground. The detailed working of each step in our methodology has been explained in upcoming sections. The proposed methodology is represented pictorially in Figure 1.
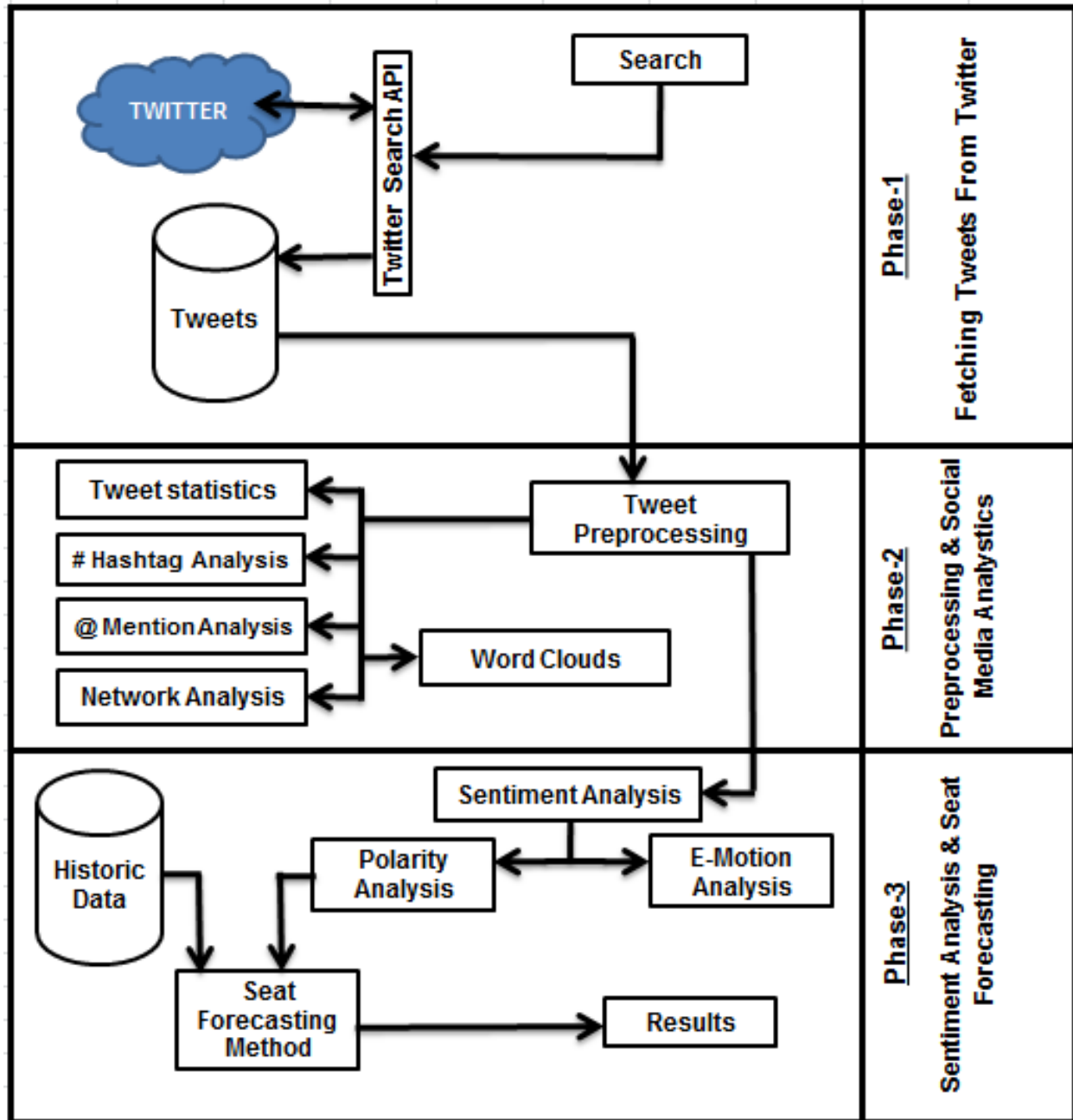
Figure 1. Proposed Methodology

6. Tweet Collection

The initial step in our proposed methodology is tweet collection from Twitter. The process of tweet collection has been done in an automated fashion. For this exercise to become fruitful, a tool has been developed using ASP.Net (Visual Studio, 2012) and integrated tweetinvi Twitter API (Application programming interface) (Tweetinvi API, 2017) for fetching tweets from Twitter. Figure 2 shows the tool developed by us for performing tweet collection task. Tweets

were fetched based on specific hashtags (#) these included party and/or popular leader's names, as shown in Table 1. Tweets were exclusively collected from Punjab state only.
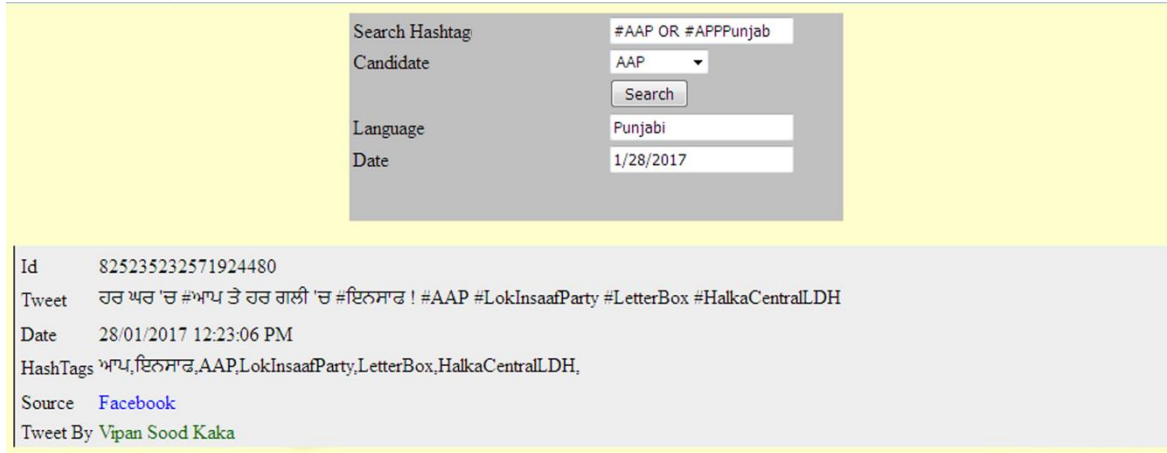
Figure 2. Tool developed for tweet collection task (Punjabi tweet)

Table 1. Hastags (#) used for fetching tweets

	SAD & BJP	**INC**	**AAP & LIP**
Hashtags (#)	#Badal, #Akali, #SAD, #BJPPunjab, #PunjabBJP, #Sukhbir, #Modi	#PunjabCongress, #Congress, #Amarinder, #Sidhu, #INC, #Captain	#AAP, #LIP, #AAPPunjab, #BhagwantMann, #Arvindkejriwal

ECI declared Punjab assembly elections for February 4, 2017. Hence, we commenced our tweet collection almost a month earlier from January 7, 2017 and this exercise continued until February 3, 2017, a day before actual voting was scheduled to take place. During this period a total of 9,157 tweets were collected in English as well as Punjabi language which is the local language of Punjab. The daily tweet collection for all the three parties is shown in Table 2.

Table 2. Daily Tweet Collection

Dates	AAP & LIP	INC	SAD & BJP
07-01-17	103	89	69
08-01-17	74	65	31
09-01-17	202	124	43
10-01-17	113	93	37
11-01-17	138	140	39
12-01-17	109	85	35
13-01-17	62	41	28
14-01-17	52	43	20
15-01-17	151	179	82
16-01-17	83	140	54
17-01-17	222	212	68
18-01-17	151	156	50

19-01-17	186	124	101
20-01-17	70	78	41
21-01-17	54	81	19
22-01-17	139	150	52
23-01-17	185	158	44
24-01-17	207	150	86
25-01-17	108	96	86
26-01-17	252	105	101
27-01-17	209	149	60
28-01-17	152	101	50
29-01-17	216	177	68
30-01-17	265	188	64
31-01-17	209	121	68
01-02-17	125	110	58
02-02-17	194	81	101
03-02-17	186	87	62
Total	**4217**	**3323**	**1617**
Percentage (%)	**46.05**	**36.28**	**17.65**

7. Data Preprocessing

Data preprocessing has often been a neglected task but it actually forms the most important constituent in the data mining process. Analyzing a set of data which has not been preprocessed could lead to conflicting results (García et al., 2015; Liu et al., 2015). Therefore, preparing data using data preprocessing approach is the most prioritized effort before executing an analysis. Since tweets, generally contains a lot of noise (unwanted stuff) (Haddi et al., 2013), wrong spellings & slang words which can ultimately affect the results (Van den Broeck et al., 2005), therefore data preprocessing is required to overcome these issues.

The task of preprocessing was performed using R-Language (R-Language, 2018). Firstly, in order to maintain normality, we removed 564 tweets from a total of 9,157 tweets which were not directly related to Punjab assembly elections. After removal of 564 tweets we were left with 8,593 tweets (AAP-LIP: 3,971, INC: 3,149, SAD-BJP: 1,473), out of which 113 tweets were in Punjabi Language. Using similar approach as used by Gaspar et al. (2016), a team consisting of an expert of Punjabi language and an expert of English language performed the transformation task of converting Punjabi language tweets into English language tweets. In order to check that no semantic meaning got lost during the conversion, two separate experts who knew both English and Punjabi languages re-verified the converted tweets. The results of transformation process are shown in Table 3.

Table 3. Result of Tweet Transformation Task

Tweet before Translation	Tweet after Translation
ਲੁਧਿਆਣਾ ਸੈਂਟਰਲ ਦੀ ਹਰ ਗਲੀ ਵਿੱਚ ਇੱਕ ਹੀ ਆਵਾਜ਼ ਗੂੰਜ ਰਹੀ ਹੈ I "ਇੱਕਲਾਬ ਜ਼ਿੰਦਾਬਾਦ" #AAP #LokInsaafParty #Letterbox	There is only one voice in every street of the Ludhiana Central, "Inquilab Zindabad". #AAP #LokInsaafParty #Letterbox

After this all the tweets were converted into lower case. Then, all the punctuations and numeric values present in the tweets were removed. Similarly, stop words (English Language) & web links were removed and finally the extra blank spaces (whitespaces) were eliminated. The end results of this phase were saved in a separate CSV file for further analysis. The final results of this phase are shown in Table 4.

Table 4. Result of Preprocessing Task

Original Tweet	#AAPkaJunoon #AAP has definitely involved more common ppl in politics then ever bfr...#AAP a much needed freshness… https://t.co/Io0MNdZUHV
Conversion to Lower Case	#aapkajunoon #aap has definitely involved more common ppl in politics then ever bfr...#aap a much needed freshness… https://t.co/io0mndzuhv
Removal of Punctuations	aapkajunoon aap definitely involved common ppl politics ever bfr aap much needed freshness httpstcoiomndzuhv
Removal of Stop Words	aap definitely involved common ppl politics ever bfr aap much needed freshness httpstcoiomndzuhv
Removal of Web Links	aapkajunoon aap definitely involved common ppl politics ever bfr aap much needed freshness
Removal of Extra white spaces	aapkajunoon aap definitely involved common ppl politics ever bfr aap much needed freshness

8. Social Media Analytics

Social media analytics is a branch of social analytics that deals with extracting important information from the data gathered from the social media websites and uses this important information for decision making. In this section, we have applied various social media analytics techniques (Chang, 2017; Chatfield et. al., 2015; Eom et. al., 2018; Stieglitz and Dang-Xuan, 2013), in order to do in-depth analysis of 2017 Punjab assembly elections.

8.1. Tweet Statistics

Tweet statistics (Purohit et al., 2013), broadly includes type of tweets, number of tweets and number of users. A total of 8,593 tweets were generated from 7,798 unique users. Among 9,157 tweets, 113 tweets were in Punjabi Language. AAP+LIP had major chunk of total tweets i.e. 4,217 (46.05%). The detailed results of tweet statistics are shown in Table 5.

Table 5. Tweet Statistics

	AAP & LIP	INC	SAD & BJP
Total Tweets	3971	3149	1473
Total Tweets (English)	3904	3116	1591
Total Tweets (Punjabi)	67	33	13
Average Tweets per day	141.82	112.46	52.60
Total Unique Senders	3589	2814	1395
Average Tweets per Sender	1.106	1.119	1.055

8.2. (#) Hashtag Analysis

Hashtag Analysis (Chae, 2015), consists of various statistics related to hashtags present in the captured tweets. A total of 641 unique (#) hashtags were identified in 8,593 tweets. Approximately 61% (5241) of tweets contained more than one hashtag. Our analysis showed that #aap was most popular hashtag with a total of 3,137 instances. The top 20 hashtags that had maximum occurrences in our tweets are shown in Figure 3.
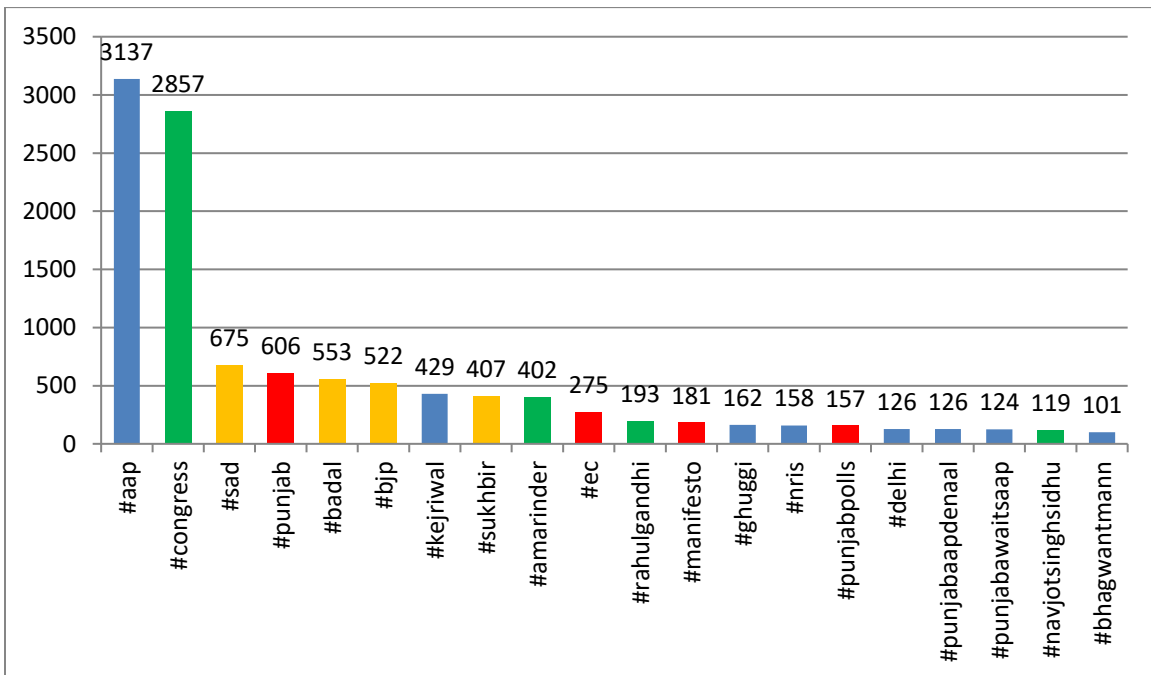

Figure 3. Top 20 (#) Hashtags with Maximum Occurrence

Blue bars (#aap, #kejriwal, #ghuggi, #nris, #delhi, #punjabaapdenaal, #bhagwantmann) in Figure 2 represent hashtags related to AAP-LIP, yellow bars (#sad, #badal, #bjp, #sukhbir) represents hashtags related to SAD-BJP, while green bars (#congress, #amarinder, #rahulgandhi, #navjotsinghsidhu) represents hashtags related to congress party. In addition to these #Punjab, #ec, #manifesto, #punjabpolls were the common hashtags represented by red bars.

8.3. @ Mention Analysis

@Mention Analysis (Shuai et al., 2012), consists of various statistics related to @mentions that were used by users. A total of 1,785 @mentions were present, out of which 874 were unique. @arvindkejriwal had maximum occurrences of 259 according to analysis. The top 10 @mentions that had maximum occurrences in our tweets are shown in Figure 4.
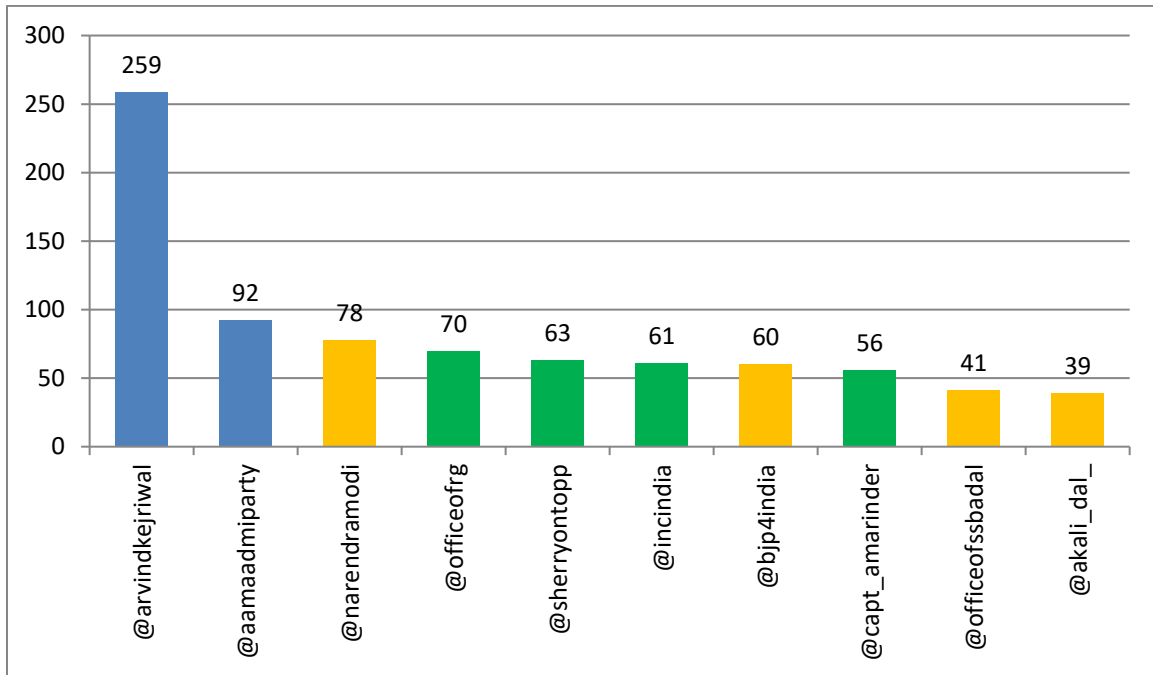
Figure 4. Top 10 @ Mentions with Maximum Occurrence

Blue bars (@arvindkejriwal, @aamaadmiparty) in Figure 3 represent @mentions related to AAP-LIP, yellow bars (@officeofrg, @sherryontopp, @incindia, @capt_amarinder) represents @mentions related to SAD-BJP, while green bars (@narendramodi, @bjp4india, @officeofssbadal, @akali_dal_) represents @mentions related to INC party.

8.4. Network Analysis

Network Analysis (HerdaĞdelen et al., 2013), consist of various communities taking active part in a particular discussion. Figure 5 shows tweet network graph for all the three political parties under consideration. It is developed using open graph visualizing platform Gephi (Bastian et al., 2009). Blue colored nodes and edges represent the tweet network for AAP-LIP, which is the largest among all political parties. Green colored nodes and edges represent the tweet network for INC. While orange colored nodes and edges represent the tweet network for SAD-BJP, which is the smallest among all political parties.

Notably, AAP-LIP clearly received the higher proportion of tweet count with 3,971 (46.21%) tweets, while INC has 3,149 (36.64%) tweets and SAD-BJP remained distant third with 1,473

(17.14%) tweets. Earlier work done by various researchers in the field of election prediction using Twitter at initial stages, applied this simple technique of counting the number of tweet mentions that a party receives (Tumasjan et al., 2010), but this technique lacks reliability as higher tweet count can be on account of both because positive as well as negative tweets. Hence, we don't consider it as an ideal technique for making a truthful prediction.
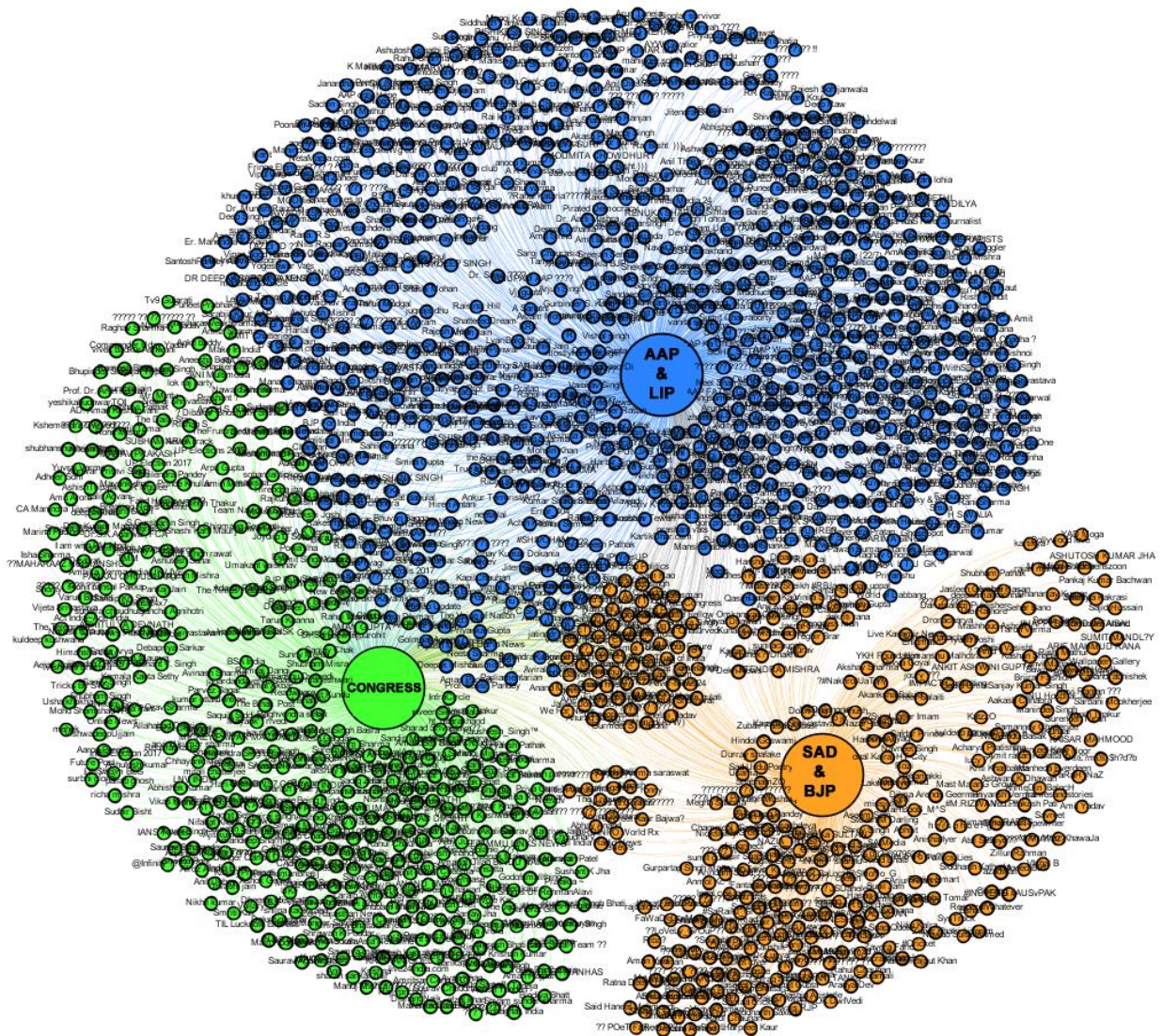

Figure 5. Tweet network graph for political parties

8.5. Word Cloud

Word cloud is a visual representation of text containing words with the highest frequencies (McNaught, and Lam, 2010). Larger the word more is its frequency in the given text. For our

analysis, we set a minimum frequency of 25. Under this criterion 250 words were selected with word "aap" having maximum frequency of 4,998 followed by word "congress" with frequency of 3,670. The corresponding word cloud of our analysis is shown in Figure 6.
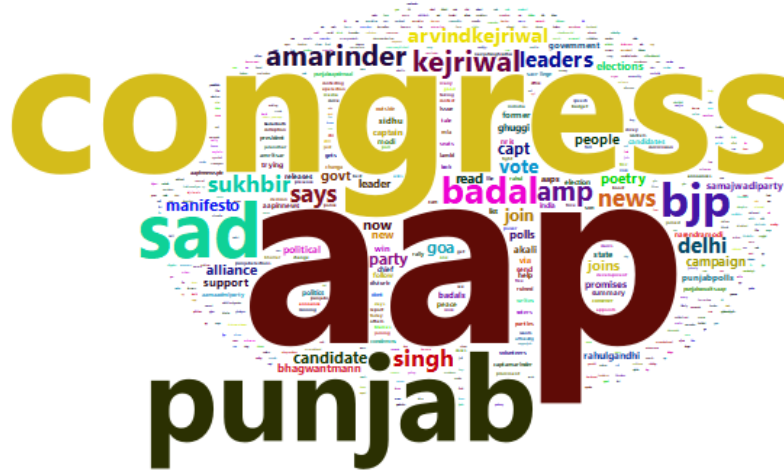

Figure 6. Word Cloud of frequent words

8.6. Sentiment Analysis

Sentiment analysis is the study of analyzing people's opinions, sentiments, evaluations, appraisals, attitudes and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. Sentiment analysis is also referred as opinion mining, opinion extraction and sentiment mining (Aswani et al., 2018; Capriotti and Ruesja, 2018; Grover et al., 2018; Liu, 2012; Mishra and Singh, 2016; Rathore and Ilavarasan, 2020; Terán and Mancera, 2019; Wu and Shen, 2015; Zavattaro et al., 2015). Under sentiment analysis, we performed two different operations (a) E-Motion Analysis (Mohammad and Turney, 2010; Ou et al., 2014) and (b) Polarity Analysis (Saif et al., 2013; Yaun et al., 2017).

(a) **E-Motion Analysis:** It is a sentiment analysis operation, in which data is categorized into eight emotions (Trust, Surprise, Sadness, Joy, Fear, Disgust, Anticipation and Anger). For our analysis we have performed the entire experimentation using syuzhet package in R-language. Figure 7 shows the results of e-motion analysis. INC had maximum trust and anticipation, while maximum values were recorded by AAP-LIP for remaining six e-motions.
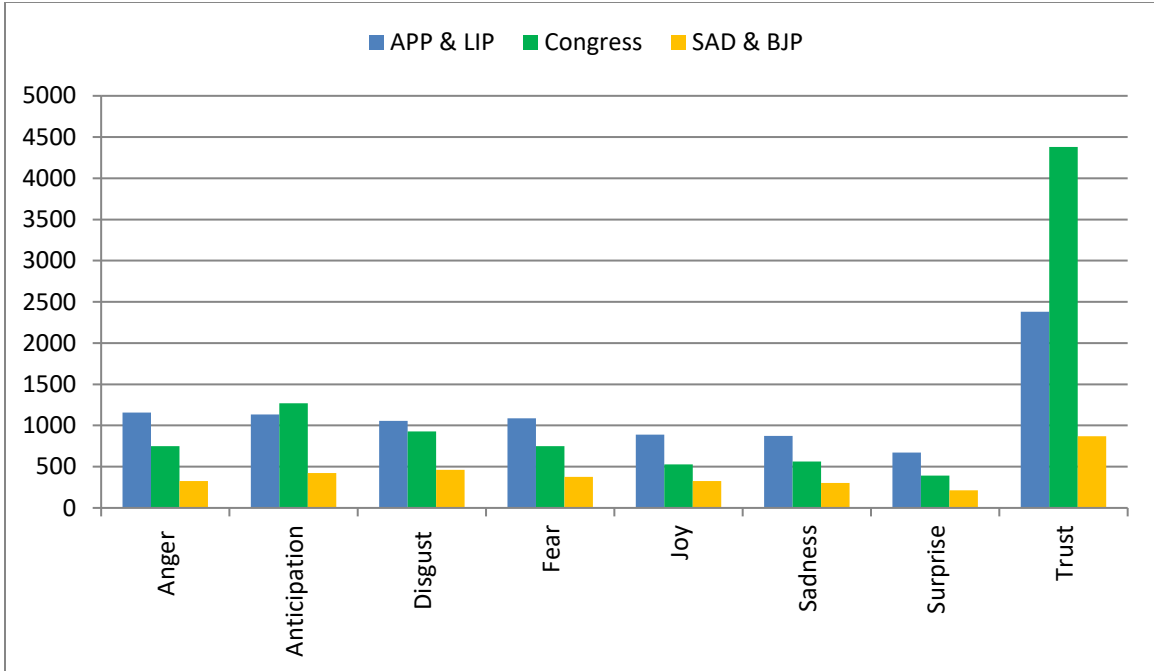
Figure 7. Results of E-Motion Analysis

(b) Polarity Analysis: Polarity analysis is a process of classifying the given text into positive or negative. For performing polarity analysis, we have used WEKA 3.8., which is an open source software and consists of a collection of machine learning algorithms for data mining tasks (Witten et al., 2016). We have used the supervised machine learning technique to accomplish the task of polarity analysis. In support of our approach, we have used two training dataset, reviews from Amazon (McAuley and Leskovec, 2013) and IMDb (Maas et al., 2011). In aggregate, dataset consists of 2000 sentences and corresponding sentiment of each sentence in form of "0" (negative) and "1" (positive). In addition to this, to achieve the higher level of efficiency we have manually annotated 1000 tweets (500 Positive and 500 Negative) collected by us. This helped us to add domain specific training data to our training dataset. Table 6 shows the statistics of our training dataset, while Table 7 shows example of the training dataset.

Table 6. Training Dataset Statistics

Source	Positive Instances	Negative Instances
Amazon	500	500
IMDb	500	500
Manually annotated tweets	500	500
Total	*1500*	*1500*

Table 7. Training Dataset Examples

Source	Positive	Negative
Amazon	So Far So Good!	I bought it for my mother and she had a problem with the battery
IMDb	The success of the film depends on the casting of Sydney Greenstreet as the Alexander Yardley character	DELETE this film from your mind!
Manually annotated tweets	a captain will beat a general in patiala: capt amarinder on jjsingh	sad violating model code of conduct: aap hr wing head navkiransingh

Further, we have built classification models using Support Vector Machine (SVM) (Schölkopf et al., 1999), K-nearest neighbor (KNN) (Aha et al., 1991) and Decision Tree (DT) (Quinlan, 1993). The aim of choosing multiple models were to check which model gives us the best results for our training data. The results of various classification models are shown in Figure 8. Since SVM outperforms the other classification models, we have used it for further operations. Figure 9 shows detailed results along with confusion matrix for SVM classification model. Further, we have used filtered classifiers, which enable us to build a classifier with a filter of our choice. SVM is used as a classifier while "StringToWordVector" is used as a filter which converts a string attribute to a vector that represents the word occurrence frequencies (Witten et al., 2016). We have employed 10-fold cross validation, also known as rotation estimation to analyze how a predictive model would perform on an unknown dataset.
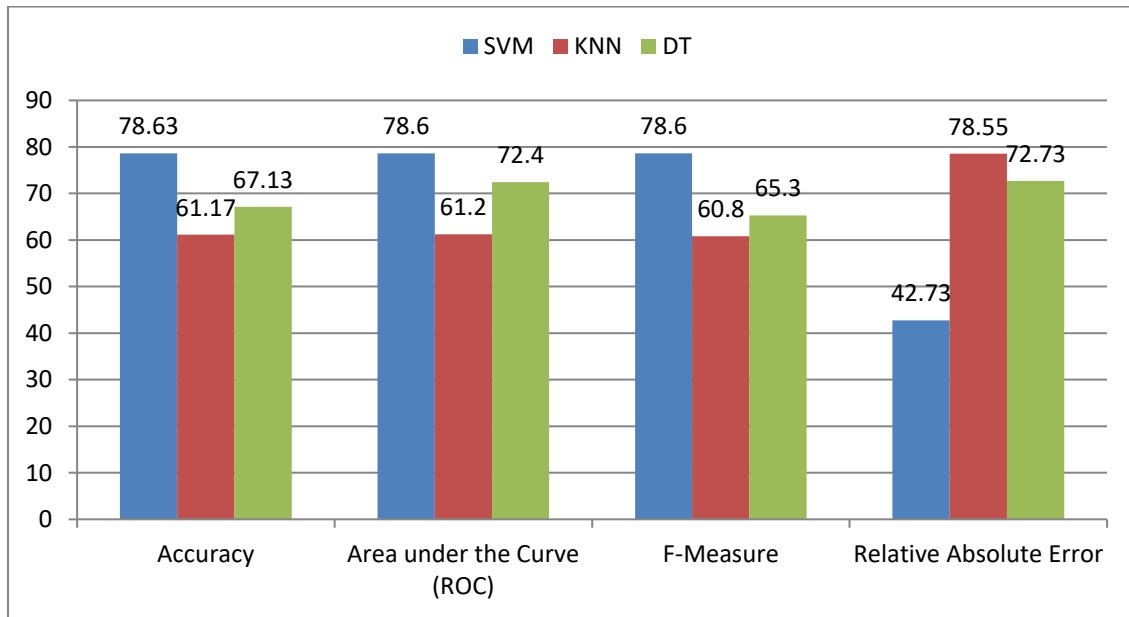
Figure 8. Results of various Classification Models

```
Time taken to build model: 22.38 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         2359               78.6333 %
Incorrectly Classified Instances        641               21.3667 %
Kappa statistic                          0.5727
Mean absolute error                      0.2137
Root mean squared error                  0.4622
Relative absolute error                 42.7333 %
Root relative squared error             92.4482 %
Total Number of Instances             3000

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC     ROC Area  PRC Area  Class
               0.795    0.223    0.781      0.795   0.788      0.573   0.786     0.724     Neg
               0.777    0.205    0.792      0.777   0.784      0.573   0.786     0.727     Pos
Weighted Avg.  0.786    0.214    0.786      0.786   0.786      0.573   0.786     0.725

=== Confusion Matrix ===

    a    b   <-- classified as
 1193  307 |   a = Neg
  334 1166 |   b = Pos
```

Figure 9. Results of Classification Model (SVM)

For testing dataset, we have used the preprocessed tweets (See section 7). Then we have exposed the testing dataset to our classification model developed using SVM and eventually we have collected classification results i.e. polarity of each tweet. Some of the instances of the results obtained after classification are shown in Table 8. Figure 10 shows the graphical representation of results obtained from the sentiment analysis.

Table 8. Result after classification of testing dataset

Source	Positive	Negative
SAD+BJP	akali will win this election with thumping majority sukhbir badal	were they sleeping for 10 years, asks capt amarinder rejecting sad manifesto promise of farm loan waiver as gim
INC	congress goes full throttle with rallies, roadshows, bike rallies on last day of campaigning in punjab	congress permanently expels 9 more dissidents from primary membership after they did not withdraw from the punjabpoll race
AAP+LIP	nris campaign for aap despite inclement weather in punjab	families of punjab farmers who committed suicides had to wait for compensation aap
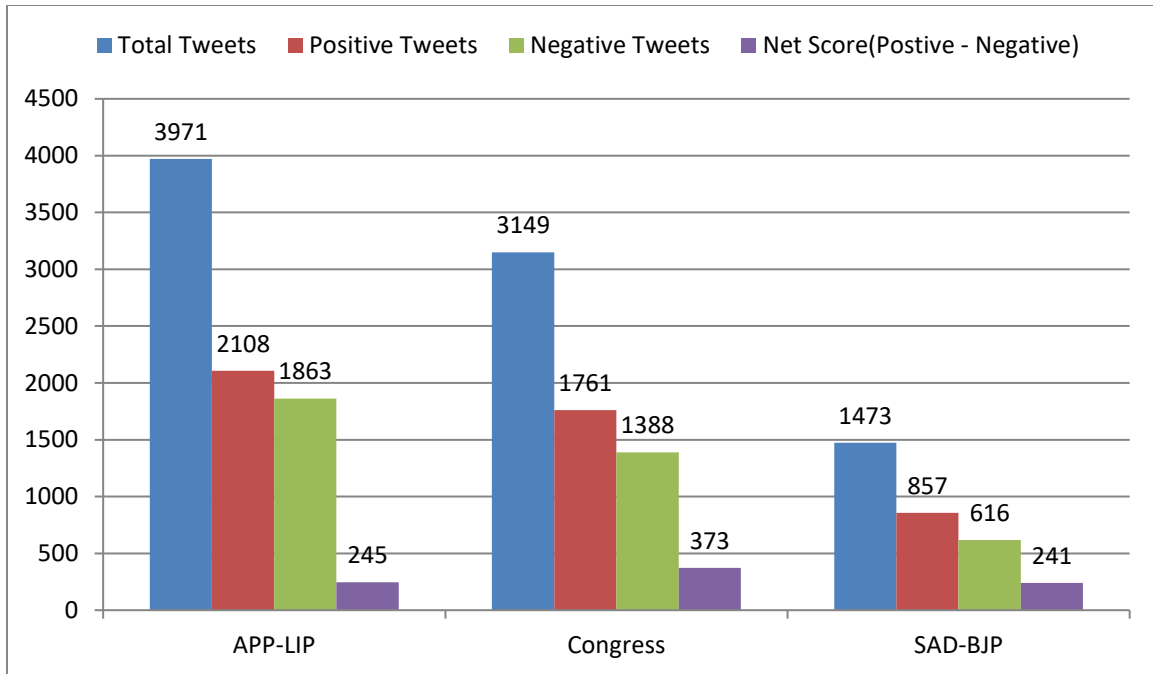
Figure 10. Results of Sentiment Analysis

Although AAP-LIP had highest tweet count but INC has the best net score (positive tweets – negative tweets). This proves our hypothesis that simply relying on the number of tweets is not a reliable way of performing the prediction task, and hence the polarity of tweets should be taken into account for performing prediction task.

9. Seat Winning Prediction

In this section, we have proposed a forecasting method using which we can predict the possible number of seats that a political party could likely win in the elections. Unlike the previous seat prediction methods (Burnap et al. 2016; Srivastava et al. 2015), our method does not give an absolute value but indeed a range under which the seats of a political party may lie. The main aim behind adopting this range criterion is to make it universal seat forecasting method that can be applied to any country or state for forecasting. Further, it is a highly challenging task to predict the exact number of seats that a political party is likely to win in the upcoming elections. The working of the seat forecasting method is explained using Algorithm 2.

Algorithm 2
Start
1: Calculate actual sentiment score (ASS) for each political party, based on the polarity of tweets.
2: Identify the political party with the highest ASS.
3: Get the historic data and rearrange the data in increasing order, on the basis of vote share (%).
4: Split the historic data into 2 subsets:
4.1: All the vote shares (%) that are less than or equal to ASS (Highest) should be put in set-1, and name it as lower bound (LB).
4.2: All the vote shares greater than ASS (Highest) should be put in set-2, and name it upper bound (UB).
5: Apply Linear Regression for seat prediction on both subsets:
5.1: Apply Linear Regression on LB, and calculate seats (S_{LB}).
5.2: Apply Linear Regression on UB, and calculate seats (S_{UB}).
6: The predicted seats are in the range of S_{LB} to S_{UB}.
End

The proposed seat forecasting method, starts with calculation of actual sentiment score (ASS) for all the political parties and the party with highest ASS is selected. Next, we have collected historic data based upon the previous Punjab assembly elections. The collected historic data has been rearranged in increasing order of vote share (%). All the historic vote shares (%) that are less than or equal to ASS (Highest) are grouped together in the set lower bound (LB). Similarly, all the historic vote shares (%) that are greater than ASS (Highest) are grouped together in the set upper bound (UB). Finally, linear regression is applied to both LB & UB to calculate the possible seat range S_{LB} to S_{UB}. The detailed working of the seat forecasting method has been explained in following sub sections.

9.1. Actual sentiment score (ASS)

This score is based upon the polarity (Positive or Negative) of tweets (See section 8.6). The positive tweets are used to calculate positive sentiment score (PSS), while the negative tweets are used to calculate negative sentiment score (NSS). Based upon PSS and NSS, ASS is calculated using the Equation 1. The ASS for each party is shown in Table 9.

$$PSS = \text{Total Number of positive tweets for a party P}$$

$$NSS = \text{Total Number of negative tweets for a party P}$$

$$ASS = \frac{(PSS \text{ of party P}) - (NSS \text{ of party P})}{\sum(PSS \text{ for all parties}) - \sum(NSS \text{ for all parties})} \; X \; 100 \tag{1}$$

Table 9. Daily Tweet Collection

	AAP & LIP	INC	SAD & BJP
Total Tweets	3971	3149	1473
PSS	2108	1761	857
NSS	1863	1388	616
PSS-NSS	245	373	241
ASS (%)	*28.52*	*43.42*	*28.05*

9.2. Historic Data

The seat forecasting method proposed by us works on the basis of historical data i.e. the results of previous elections. Although, the state of Punjab came into existence in the year 1966 but the number of assembly seats were not 117 at that time. Hence we have considered only that historical data in which the number of assembly seats were 117, similar to current date scenario. The details of the historic data are shown in Table 10. Since INC has the highest ASS (43.42), we have split the historic data accordingly. The vote share (%) which are less than or equal to 43.42 are placed into one subset (Lower Bound) while the remaining are placed into the other subset (Upper Bound). The historic data along with two subsections is shown in Table 11.

Table 10. Historic Election Data (ECI, 2018b)

Year	Vote Share (%)	Number of Seats	Winner Political Party
2012	41.91	68	SAD & BJP
2007	45.37	67	SAD & BJP
2002	35.81	62	INC
1997	45.97	93	SAD & BJP
1992	43.83	87	INC
1985	43.00	79	SAD & BJP
1980	45.19	63	INC
1977	46.40	83	SAD & Janata Party

Table 11. Rearranged Historic Election Data

Vote Share (%)	Number of Seats	Section
35.81	62	Lower
41.91	68	Bound
43.00	79	Subset
43.42	To be Determined	
43.83	87	
45.19	63	Upper
45.37	67	Bound
45.97	93	Subset
46.4	83	

9.3. Seats Forecasting

This is the final phase of the proposed method. Based upon the lower and the upper bound the possible seat range is calculated using another supervised machine learning algorithm, linear regression (Montgomery et al., 2012) and is implemented using WEKA. Linear regression establishes a mathematical relationship between a dependent variable and an independent variable using Equation 2, where 'y' is the estimator for the dependent variable, 'a' is an intercept of 'y', 'b' is the regression coefficient and 'x' is score of independent variable.

$$y = a + bx \qquad (2)$$

In our case the independent variable is the vote share (%), and the dependent variable is the seats, as we are finding the number of seats based upon the vote share (%). The training dataset for the prediction is historic data, while the testing dataset is the ASS of INC party i.e. 43.42%. Based upon the lower bound subset values, the linear regression gave a minimum value of seats as 75.69 (~76 Seats). Similarly, for upper bound subset values the linear regression gave a maximum value of 78.6 (~79). So the predicted range lies between 76 to 79 seats for a given vote share of 43.42%. Figure 11 and Figure 12 shows scattered plots along with the best fit slope for the lower and the upper bound subset respectively. The blue color points represent the historic data while the red color point represents the predicted data.
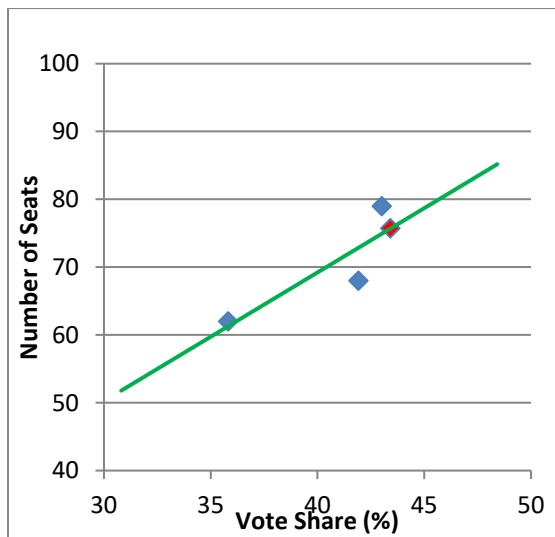

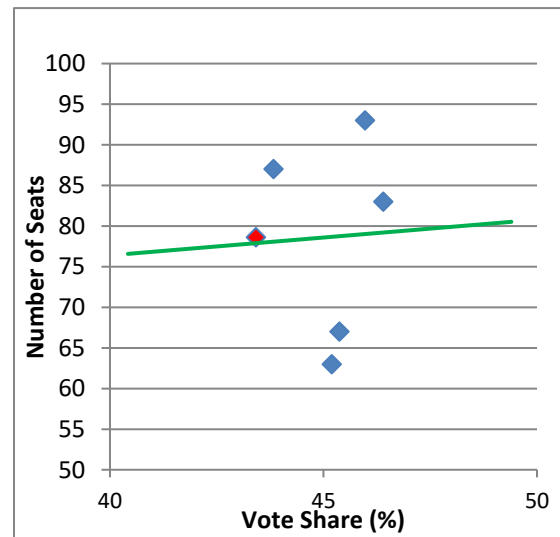**Figure 11**. Linear Regression on Lower Bound Subset

Figure 12. Linear Regression on Upper Bound Subset

According to our proposed method, the INC was likely to win 76 to 79 seats in 2017 Punjab assembly elections and they actually won 77 seats, which was well within our predicted range. For calculating the number of seats for party finishing second and third, Equations 3a, 3b, 4a and 4b are used. The details of the seats are shown in Table 12.

T=117 (Total Number of Seats), F_U=79 (Upper Seat Limit of 1st party), F_L=76 (Lower Seat Limit of 1st party), S_U= Upper Seat Limit of 2nd party, S_L= Lower Seat Limit of 2nd party, L_U= Upper Seat Limit of 3rd party, L_L= Lower Seat Limit of 3rd party.

$$S_U = \left[\frac{(T-F_L) \text{ X (ASS of 2nd) X 2}}{100} \right] \tag{3a}$$

$$S_L = \left[\frac{(T-F_U) \text{ X (ASS of 2nd) X 2}}{100} \right] \tag{3b}$$

$$L_U = (T - F_L - S_L) \tag{4a}$$

$$L_L = (T - F_U - S_U) \tag{4b}$$

Table 12. Likely Seats of Each Party

	ASS (%)	Upper Seat Limit	Lower Seat Limit	Seat Range
INC	43.42	79	76	79-76
AAP & LIP	28.52	23	22	23-22
SAD & BJP	28.05	19	15	19-15

10. Discussion

Methodologically, our research work presents a technique which can predict the outcome of an election along with number of seats from the user generated social media content (tweets). In this section, we have compared our computed results with the actual results and have further compared our method with other seat prediction methods (See Section 3). In addition to these strategies, we have discussed the implication for various stakeholders. Finally, limitations are discussed with future directions.

10.1. Comparison with Actual Results

Based upon the results of our seat forecasting method, the INC must win 76 to 79 seats. However, the results are of no importance if they are not validated and compared with actual results. The voting for 2017 Punjab Assembly Elections was held on February 4, 2017 with results declared on March 11, 2017. Our prediction for INC proved correct as they won 77 seats which exactly lie in the range (76 to 79) as predicted by us. Similarly, the ASS (43.42%) calculated by us was quite close to the actual vote share (%) of INC (38.5%) (ECI, 2017). Further, the seats won by party finishing second and third were also exactly lying in the range as predicted by us. The details of the same along with MAE are shown in Table 13.

Table 13. Comparison of Results (ECI, 2017)

	Predicted Results		Actual Results		MAE %
	ASS (%)	Seats	Vote Share (%)	Seats	(ASS – Vote Share)
INC	43.42	79-76	38.5	77	+4.92
AAP & LIP	28.52	23-22	24.9	22	+3.62
SAD & BJP	28.05	19-15	30.6	18	-2.55

[**Note:** Vote share of AAP & LIP was 23.7 and 1.2 respectively. Similarly vote share of SAD & BJP was 25.2 and 5.4 respectively]

However, there does exist a difference between the actual vote share and ASS calculated by us. One of the possible reasons for this divergence can be the fact that on many assembly seats the candidates finishing first and second were separated by just few hundred votes. Further, we have performed the entire experimentation for INC, AAP-LIP and SAD-BJP only but there were many other parties and independent candidates who though didn't win yet drew a large volume of votes (ECI, 2017) and led to this contradiction.

10.2. Comparison with Other Seat Forecasting Techniques

After an intensive literature review, we came across only two papers i.e. Srivastava et al. (2015) and Burnap et al. (2016) that have used any Twitter based seat forecasting method. Srivastava et al. (2015) predicted the results of 2015 Delhi assembly elections. They calculated positive sentiment share for each party based upon sentiment analysis. Further they used mapping functions to convert this positive sentiment share into seat share. Their results achieved high degree of accuracy, setting a benchmark for other techniques.

Burnap et al. (2016) predicted the seats for 2015 UK general election. They used sentiment analysis on Twitter data to calculate positive party sentiment for each party. Based upon this positive party sentiment, seat forecasting was done on constituency by constituency (seat by seat) basis by applying national swing on UK 2010 general election results. However the calculated results proved to be incorrect, making it not suitable for future adoption.

Though both of these methods were novel and served as a major motivation for our research work. However, both of these techniques suffered from two major short comings, firstly negative tweets or negative sentiment scores were not considered (Burnap et al., 2016 considered partial negative scores while ignoring majority of them) and secondly, both methods were trying to predict the exact number of seats. Since prediction is a trivial task and it is highly impossible to precisely predict the exact number of seats a party will be winning in the elections. Hence, to overcome these problems, we have considered both positive as well as negative tweets for our calculations. Further, we have proposed a seat forecasting method that gives us a range of seats instead of exact number that a party is likely to win in the elections. Using the range method expands the scope of correct predictions, making it better than the previously discussed methods.

10.3. Implications

The implication of this study for practice has been divided into three subsections (a) Political Parties contesting the elections (b) General Public which votes to elect new leaders and (c) Opinion poll experts. These subsections are discussed briefly below.

(a) Politicians/ Political Parties: Since political parties and their party heads are continuously engaged in social media for communication with public and ultimately these people are the one who are going to vote to elect their leaders. They can see which leader is more popular in certain region or constituency so that he can be given chance to contest from that particular region or constituency, so that there is more chance of a victory. Further, parties can monitor what are the demands of people so that they can add that in their manifesto, which will again be beneficial for these political parties.

(b) People: They are the principal stake holders, as ultimately they are the ones who actually vote and elect the new government. Having direct contact with the political leaders is really beneficial and people do prefer electing those leaders who are easily accessible. As literature survey suggests the social media engagement increases the chances of politicians of winning an election. Further, people can get insights that which leader is promising what and how he/she has performed in his/her previous tenure or he/she is a novice in politics. All this really guide the general public to elect their leader wisely.

(c) Opinion poll experts: Generally, opinion poll is biased as it contains only a certain number of people whose responses were taken and that too from a particular region or constituency. But using this technique, opinion poll expert can give an unbiased opinion, as social media data is gathered from different people covering different regions. Most elections in India are held in phased manner, so opinion polls of earlier phases many times do affect the voting of other phases. Hence, this technique can be used to give unbiased opinion polls. Further, we can improve the preciseness of the opinion polls by combining the proposed technique with survey approach.

10.4 Limitation and Future Research Directions

Though, our study has given us very encouraging results and it is clearly depicted that our proposed technique can be applied as a generic method to predict the election outcomes based upon Twitter data. However, our study suffers from a few limitations and these studies can certainly be addressed in the future so that we can get more precise and accurate results. The main limitation of this study is dataset of 9,157 tweets, on which the entire experimentation has been performed. As the tweet collection period considered by us is approximately one month, definitely we can expand this period to at least six months or more. Further, if tweets could be collected for a larger timeline, we can check what all events lead to switching of loyalties from

one political party to another. We can also check the monthly timelines, which will give us clear indication where a particular party lost its momentum when compared to its counter parts. Similarly, bot detection (Chu et al. 2010) has also not performed in our proposed technique. Social media bots can adversely affect the prediction, by making the results biased. Hence, it is very important to address them. Further, only original tweets have been considered for analysis, while no emphasis has been given to re-tweets. Re-tweets as we know constitute up to third of the entire Twitter traffic (Holmberg, 2014), hence considerable amount of Twitter traffic has been ignored by us in our analysis. In addition to this, many social media analytics techniques like topic modeling and location based analysis has not been used. These techniques would have given better results and should be included in future studies. Moreover, in our technique only Twitter data has been used, however future studies can also incorporate data from other social media platforms like Instagram. This will open whole new horizons for researchers as concept of digital image processing (Pitas, 2002) can be used to fetch details from the images.

11. Conclusions

In recent times Twitter has been extensively used for election prediction. However, none of the work done till date to the best of our knowledge has been able to provide a generic technique that could predict the outcome of any election taking place anywhere in the world based upon the Twitter data. The aim of this paper was to highlight the success of our proposed approach to predict the winner of 2017 Punjab Assembly Elections along with conformity about the percentage vote share and hence the number of seats that each party was likely to get. Since prediction of a future event has always been a trivial task and it is impracticable to predict the exact number of seats. Therefore our seat forecasting method contains the range (Upper and lower limit) that a party is likely to win based on historical data.

According to our research INC was predicted to win the 2017 Punjab Assembly Elections by getting a total of 79-76 seats from the available 117, which proved correct as they won 77 seats. Not only that we also correctly predicted the seats won by political parties finishing second (AAP-LIP) and third (SAD-BJP). This clearly shows that the proposed method of election prediction using historic data and regression can be considered as a reliable technique to predict the results of future elections. Further, this proposed method is not only for Punjab Assembly Elections, but can be used as a generic solutions for all the countries/states etc where the winning political party is the one which has a maximum number of seats.

However, we perceive that there is a scope of further improvement by refining this research work to a higher step by introducing the concept of location based tweet analysis. This would enable us to make prediction at district level or even precisely for a particular assembly seat.

References

Abascal-Mena, R., Lema, R., & Sèdes, F. (2015). Detecting sociosemantic communities by applying social network analysis in tweets. *Social Network Analysis and Mining*, *5*(1), 38.

ABP News (2016) [Online], http://abpnews.abplive.in/india-news/prashant-kishor-to-help-congress-in-punjab-assembly-elections-330811 (Accessed on January 15, 2018).

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine learning*, *6*(1), 37-66. DOI: 10.1007/BF00153759

Ahmed, S., Jaidka, K., & Cho, J. (2016). The 2014 Indian elections on Twitter: A comparison of campaign strategies of political parties. Telematics and Informatics, 33(4), 1071-1087.

Alalwan, A. A., Rana, N. P., Dwivedi, Y. K., & Algharabat, R. (2017). Social media in marketing: A review and analysis of the existing literature. *Telematics and Informatics*, 34(7), 1177-1190.

Aladwani, A. M., & Dwivedi, Y. K. (2018). Towards a theory of SocioCitizenry: Quality anticipation, trust configuration, and approved adaptation of governmental social media. *International Journal of Information Management*, 43, 261-272.

Aswani, R., Kar, A. K., Ilavarasan, P. V., & Dwivedi, Y. K. (2018). Search engine marketing is not all gold: Insights from Twitter and SEOClerks. *International Journal of Information Management*, 38(1), 107-116.

Attu, R., & Terras, M. (2017). What people study when they study Tumblr: Classifying Tumblr-related academic research. *Journal of Documentation*, *73*(3), 528-554.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *Icwsm*, *8*, 361-362.

Bertot, J. C., Jaeger, P. T., & Hansen, D. (2012). The impact of polices on government social media usage: Issues, challenges, and recommendations. *Government information quarterly*, *29*(1), 30-40.

Blazquez, D., & Domenech, J. (2017). Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*. DOI: 10.1016/j.techfore.2017.07.027

Boynton, G. R., & Richardson Jr, G. W. (2016). Agenda setting in the twenty-first century. *New Media & Society*, 18(9), 1916-1934.

Bruns, A., & Stieglitz, S. (2013). Towards more systematic Twitter analysis: metrics for tweeting activities. *International Journal of Social Research Methodology*, *16*(2), 91-108.

Burnap, P., Gibson, R., Sloan, L., Southern, R., & Williams, M. (2016). 140 characters to victory?: Using Twitter to predict the UK 2015 General Election. *Electoral Studies*, *41*, 230-233. DOI: 10.1016/j.electstud.2015.11.017

Business Standard (2016) [Online], http://www.business-standard.com/article/politics/aap-suffers-a-blow-after-infighting-in-punjab-116082900802_1.html (Accessed on January 15, 2018).

Capriotti, P., & Ruesja, L. (2018). How CEOs use Twitter: A comparative analysis of Global and Latin American companies. *International Journal of Information Management*, 39, 242-248.

Ceron, A., Curini, L., Iacus, S. M., & Porro, G. (2014). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with

an application to Italy and France. *New Media & Society*, *16*(2), 340-358. DOI: 10.1177/1461444813480466

Chae, B. K. (2015). Insights from hashtag# supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics*, *165*, 247-259. DOI: 10.1016/j.ijpe.2014.12.037

Chang, V. (2017). A proposed social network analysis platform for big data analytics. *Technological Forecasting and Social Change*. DOI: 10.1016/j.techfore.2017.11.002

Chatfield, A. T., Reddick, C. G., & Brajawidagda, U. (2015). Government surveillance disclosures, bilateral trust and Indonesia–Australia cross-border security cooperation: Social network analysis of Twitter data. *Government Information Quarterly*, *32*(2), 118-128.

Chatfield, A. T., Scholl, H. J. J., & Brajawidagda, U. (2013). Tsunami early warnings via Twitter in government: Net-savvy citizens' co-production of time-critical public information services. Government information quarterly, 30(4), 377-386.

Cody, E. M., Reagan, A. J., Mitchell, L., Dodds, P. S., & Danforth, C. M. (2015). Climate change sentiment on Twitter: An unsolicited public opinion poll. PloS one, 10(8), e0136092.

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2010, December). Who is tweeting on Twitter: human, bot, or cyborg?. In Proceedings of the 26th annual computer security applications conference (pp. 21-30). ACM.

DiGrazia, J., McKelvey, K., Bollen, J., & Rojas, F. (2013). More tweets, more votes: Social media as a quantitative indicator of political behavior. *PloS one*, *8*(11), e79449. DOI: 10.1371/journal.pone.0079449

Dwivedi, Y. K., Kapoor, K. K., & Chen, H. (2015). Social media marketing and advertising. *The Marketing Review*, 15(3), 289-309.

Domingo, J., & Martos, J. M. (2016). Analysis of the political discourse in Spain about school failure on Twitter. education policy analysis archives, 24, 70.

ECI (2014), 2014 General Election Report [Online], http://eci.nic.in/eci_main/archiveofge2014/4%20-%20List%20of%20Successful%20Candidates.pdf (Accessed on January 15, 2018).

ECI (2017), Punjab Assembly Election Results 2017 [Online], "http://eci.nic.in/eci_main/archiveofge2017/Punjab/05.Performance%20of%20Political%20Parties.pdf". (Accessed January 10, 2018).

ECI (2018a), About ECI [Online], http://eci.nic.in/eci_main1/the_setup.aspx (Accessed on January 15, 2018).

ECI (2018b), Punjab assembly election results [Online], http://eci.nic.in/eci_main1/ElectionStatistics.aspx (Accessed on January 15, 2018).

Eom, S. J., Hwang, H., & Kim, J. H. (2018). Can social media increase government responsiveness? A case study of Seoul, Korea. *Government information quarterly*, *35*(1), 109-122.

Firstpost (2016a) [Online], http://www.firstpost.com/politics/congress-captain-amarinder-singh-goes-for-a-makeover-before-punjab-assembly-polls-2017-2881210.html (Accessed on January 15, 2018).

Firstpost (2016b) [Online], http://www.firstpost.com/politics/navjot-singh-sidhu-quits-as-rajya-sabha-member-may-become-aaps-cm-candidate-for-punjab-2900866.html (Accessed on January 15, 2018).

Ganis, M., & Kohirkar, A. (2015). *Social media analytics: Techniques and insights for extracting business value out of social media*. IBM Press.

García, S., Luengo, J., & Herrera, F. (2016). *Data preprocessing in data mining*. Springer. DOI: 10.1007/978-3-319-10247-4

Gaspar, R., Pedro, C., Panagiotopoulos, P., & Seibt, B. (2016). Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events. *Computers in Human Behavior*, *56*, 179-191. DOI: 10.1016/j.chb.2015.11.040

Gayo Avello, D., Metaxas, P. T., & Mustafaraj, E. (2011). Limits of electoral predictions using twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. Association for the Advancement of Artificial Intelligence.

Gayo-Avello, D. (2012). " I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper"--A Balanced Survey on Election Prediction using Twitter Data. *arXiv preprint arXiv:1204.6441*.

Gintova, M. (2019). Understanding government social media users: an analysis of interactions on Immigration, Refugees and Citizenship Canada Twitter and Facebook. *Government Information Quarterly*. DOI: 10.1016/j.giq.2019.06.005

Grover, P., Kar, A. K., & Davies, G. (2018). "Technology enabled Health"–Insights from twitter analytics with a socio-technical perspective. *International Journal of Information Management*, 43, 85-97.

Grover, P., Kar, A. K., Dwivedi, Y. K., & Janssen, M. (2019). Polarization and acculturation in US Election 2016 outcomes–Can twitter analytics predict changes in voting preferences. *Technological Forecasting and Social Change*, 145, 438-460.

Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. *Procedia Computer Science*, *17*, 26-32. DOI: 10.1016/j.procs.2013.05.005

Hargharcaptain Website (2018) [Online], http://hargharcaptain.com/ (Accessed on January 15, 2018).

Harris, J. K., Moreland-Russell, S., Choucair, B., Mansour, R., Staub, M., & Simmons, K. (2014). Tweeting for and against public health policy: response to the Chicago Department of Public Health's electronic cigarette Twitter campaign. Journal of medical Internet research, 16(10), e238.

HerdaĞdelen, A., Zuo, W., Gard-Murray, A., & Bar-Yam, Y. (2013). An exploration of social identity: The geography and politics of news-sharing communities in twitter. *Complexity*, *19*(2), 10-20. DOI: 10.1002/cplx.21457

Hindustan Times (2017) [Online], http://www.hindustantimes.com/india-news/navjot-singh-sidhu-joins-congress-ahead-of-punjab-elections/story-jY9mpBfPX2LfxzWAvH44nI.html (Accessed on January 15, 2018).

Holmberg, K. (2014). The impact of retweeting on altmetrics. Retrieved July 8, 2014, from http://de.slideshare.net/kholmber/the-meaning-of-retweeting.

Howard, P. N., & Parks, M. R. (2012). Social media and political change: Capacity, constraint, and consequence. *Journal of communication*, *62*(2), 359-362. DOI: 10.1111/j.1460-2466.2012.01626.x

Hutchins, B. (2016). The many modalities of social networking: The role of Twitter in greens politics. Environmental Communication, 10(1), 25-42.

Ibrahim, N. F., Wang, X., & Bourne, H. (2017). Exploring the effect of user engagement in online brand communities: Evidence from Twitter. Computers in Human Behavior, 72, 321-338.

Indian Express (2016) [Online], http://indianexpress.com/article/india/politics/prashant-kishor-meets-capt-amarinder-discusses-strategy-for-punjab-elections/ (Accessed on January 15, 2018).

Jungherr, A. (2013, October). Tweets and votes, a special relationship: the 2009 federal election in germany. In *Proceedings of the 2nd workshop on Politics, elections and data* (pp. 5-14). ACM. DOI: 10.1145/2508436.2508437

Jungherr, A., Jürgens, P., & Schoen, H. (2012). Why the pirate party won the german election of 2009 or the trouble with predictions: A response to tumasjan, a., sprenger, to, sander, pg, & welpe, im "predicting elections with twitter: What 140 characters reveal about political sentiment". *Social science computer review*, *30*(2), 229-234. DOI: 10.1177/0894439311404119

Kamboj, S., Sarmah, B., Gupta, S., & Dwivedi, Y. (2018). Examining branding co-creation in brand communities on social media: Applying the paradigm of Stimulus-Organism-Response. *International Journal of Information Management*, 39, 169-185.

Kapoor, K. K., & Dwivedi, Y. K. (2015). Metamorphosis of Indian electoral campaigns: Modi's social media experiment. *International Journal of Indian Culture and Business Management*, *11*(4), 496-516. DOI: 10.1504/IJICBM.2015.072430

Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in social media research: past, present and future. *Information Systems Frontiers*, 20(3), 531-558.

Kassarjian, H. H. (1977). Content analysis in consumer research. *Journal of Consumer Research*, *4*(1), 8-18.

Kavanaugh, A. L., Fox, E. A., Sheetz, S. D., Yang, S., Li, L. T., Shoemaker, D. J., ... & Xie, L. (2012). Social media use by government: From the routine to the critical. *Government Information Quarterly*, *29*(4), 480-491.

Kayser, V., & Blind, K. (2017). Extending the knowledge base of foresight: The contribution of text mining. *Technological Forecasting and Social Change*, *116*, 208-215.

Kelm, O., Dohle, M., & Bernhard, U. (2017). Social media activities of political communication practitioners: The impact of strategic orientation and in-group orientation. International Journal of Strategic Communication, 11(4), 306-323.

Khan, N. A., & Khan, A. N. (2019). What followers are saying about transformational leaders fostering employee innovation via organisational learning, knowledge sharing and social media use in public organisations?. *Government Information Quarterly*. DOI: 10.1016/j.giq.2019.07.003

Kruikemeier, S., van Noort, G., Vliegenthart, R., & H. de Vreese, C. (2016). The relationship between online campaigning and political involvement. Online information review, 40(5), 673-694.

LaMarre, H. L., & Suzuki-Lambrecht, Y. (2013). Tweeting democracy? Examining Twitter as an online public relations strategy for congressional campaigns'. Public Relations Review, 39(4), 360-368.

Lawrence, E., Sides, J., & Farrell, H. (2010). Self-segregation or deliberation? Blog readership, participation, and polarization in American politics. Perspectives on Politics, 8(1), 141-157.

Lim, M. (2012). Clicks, cabs, and coffee houses: Social media and oppositional movements in Egypt, 2004–2011. *Journal of communication*, *62*(2), 231-248. DOI: 10.1111/j.1460-2466.2012.01628.x

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), 1-167. DOI: 10.2200/S00416ED1V01Y201204HLT016

Liu, Y., Chen, Y., Wu, S., Peng, G., & Lv, B. (2015). Composite leading search index: a preprocessing method of internet search data for stock trends prediction. Annals of Operations Research, 234(1), 77-94.

Llewellyn, C., Grover, C., Alex, B., Oberlander, J., & Tobin, R. (2015, September). Extracting a topic specific dataset from a Twitter archive. In *International Conference on Theory and Practice of Digital Libraries* (pp. 364-367). Springer, Cham.

Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011, June). Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1* (pp. 142-150). Association for Computational Linguistics.

Makazhanov, A., Rafiei, D., & Waqar, M. (2014). Predicting political preference of Twitter users. *Social Network Analysis and Mining*, *4*(1), 193. DOI: 10.1007/s13278-014-0193-5

McAuley, J., & Leskovec, J. (2013, October). Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems* (pp. 165-172). ACM. DOI: 10.1145/2507157.2507163

McNaught, C., & Lam, P. (2010). Using Wordle as a supplementary research tool. *The qualitative report*, *15*(3), 630.

Mellon, J. (2014). Internet search data and issue salience: The properties of Google Trends as a measure of issue salience. *Journal of Elections, Public Opinion & Parties*, *24*(1), 45-72. DOI: 10.1080/17457289.2013.846346

Mishra, N., & Singh, A. (2016). Use of twitter data for waste minimisation in beef supply chain. Annals of Operations Research, 1-23.

Mohammad, S. M., & Turney, P. D. (2010, June). Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text* (pp. 26-34). Association for Computational Linguistics.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (Vol. 821). John Wiley & Sons.

News 18 (2016) [Online], http://www.news18.com/news/politics/several-sad-aap-and-bjp-leaders-join-congress-in-punjab-1327489.html (Accessed on January 15, 2018).

Ortiz, J. R., Zhou, H., Shay, D. K., Neuzil, K. M., Fowlkes, A. L., & Goss, C. H. (2011). Monitoring influenza activity in the United States: a comparison of traditional surveillance systems with Google Flu Trends. *PloS one*, *6*(4), e18687. DOI: 10.1371/journal.pone.0018687

Ou, G., Chen, W., Wang, T., Wei, Z., Binyang, L. I., Yang, D., & Wong, K. F. (2014). Exploiting community emotion for microblog event detection. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1159-1168).

Panagiotopoulos, P., Barnett, J., Bigdeli, A. Z., & Sams, S. (2016). Social media in emergency management: Twitter as a tool for communicating risks to the public. *Technological Forecasting and Social Change*, *111*, 86-96. DOI: 10.1016/j.techfore.2016.06.010

Pitas, I. (2000). Digital image processing algorithms and applications. John Wiley & Sons.

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific reports*, *3*, srep01684. DOI: 10.1038/srep01684

Punjabi Tribune (2016) [Online], http://www.punjabtribune.com/news/print.aspx?ac=13733 (Accessed on January 15, 2018).

Purohit, H., Hampton, A., Shalin, V. L., Sheth, A. P., Flach, J., & Bhatt, S. (2013). What kind of# conversation is Twitter? Mining# psycholinguistic cues for emergency coordination. *Computers in Human Behavior*, *29*(6), 2438-2447. DOI: 10.1016/j.chb.2013.05.007

Quinlan, J. R. (1993). C4. 5: Programming for machine learning. *Morgan Kauffmann*, *38*, 48.

Rad, A. A., Jalali, M. S., & Rahmandad, H. (2017). How exposure to different opinions impacts the life cycle of social media. Annals of Operations Research, 1-29

Rathore, A. K., & Ilavarasan, P. V. (2020). Pre-and post-launch emotions in new product development: Insights from twitter analytics of three products. *International Journal of Information Management*, 50, 111-127.

Rauchfleisch, A., & Metag, J. (2016). The special case of Switzerland: Swiss politicians on Twitter. New Media & Society, 18(10), 2413-2431.

R-Language (2018) [Online], https://www.r-project.org/about.html (Accessed January 10, 2018).

Saboo, A. R., Kumar, V., & Park, I. (2016). Using Big Data to Model Time-Varying Effects for Marketing Resource (Re) Allocation. *MIS Quarterly*, *40*(4).

Safiullah, M., Pathak, P., Singh, S., & Anshul, A. (2017). Social media as an upcoming tool for political marketing effectiveness. Asia Pacific Management Review, 22(1), 10-15.

Saif, H., Fernandez, M., He, Y., & Alani, H. (2013). Evaluation datasets for Twitter sentiment analysis: a survey and a new dataset, the STS-Gold.

Sang, E. T. K., & Bos, J. (2012, April). Predicting the 2011 dutch senate election results with twitter. In *Proceedings of the workshop on semantic analysis in social media* (pp. 53-60). Association for Computational Linguistics.

Schölkopf, B., Burges, C. J., & Smola, A. J. (Eds.). (1999). *Advances in kernel methods: support vector learning*. MIT press.

Shuai, X., Pepe, A., & Bollen, J. (2012). How the scientific community reacts to newly submitted preprints: Article downloads, twitter mentions, and citations. *PloS one*, *7*(11), e47523. DOI: 10.1371/journal.pone.0047523

Singh, P., Dwivedi, Y. K., Kahlon, K. S., & Sawhney, R. S. (2018a, June). Intelligent Monitoring and Controlling of Public Policies Using Social Media and Cloud Computing. In International Working Conference on Transfer and Diffusion of IT (pp. 143-154). Springer, Cham.

Singh, P., Dwivedi, Y. K., Kahlon, K. S., Sawhney, R. S., Alalwan, A. A., & Rana, N. P. (2019). Smart Monitoring and Controlling of Government Policies Using Social Media and Cloud Computing. *Information Systems Frontiers*, 1-23. DOI: 10.1007/s10796-019-09916-y

Singh, P., Kahlon, K. S., Sawhney, R. S., Vohra, R., & Kaur, S. (2018b). Social media buzz created by# nanotechnology: insights from Twitter analytics. Nanotechnology Reviews, 7(6), 521-528.

Singh, P., Sawhney, R. S., & Kahlon, K. S. (2017, November). Forecasting the 2016 US Presidential Elections Using Sentiment Analysis. In *Conference on e-Business, e-Services and e-Society* (pp. 412-423). Springer, Cham. DOI: 10.1007/978-3-319-68557-1_36

Singh, P., Sawhney, R. S., & Kahlon, K. S. (2018c). Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government. *ICT Express*, *4*(3), 124-129.

Srivastava, R., Kumar, H., Bhatia, M. P., & Jain, S. (2015, September). Analyzing Delhi assembly election 2015 using textual content of social network. In *Proceedings of the Sixth International Conference on Computer and Communication Technology 2015* (pp. 78-85). ACM. DOI: 10.1145/2818567.2818582

Statista (2018), Global Twitter Users [Online], https://www.statista.com/statistics/303681/twitter-users-worldwide/ (Accessed on January 15, 2018).

Stephens, M., & Poorthuis, A. (2015). Follow thy neighbor: Connecting the social and the spatial networks on Twitter. *Computers, Environment and Urban Systems*, *53*, 87-95.

Stieglitz, S., & Dang-Xuan, L. (2013). Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining*, *3*(4), 1277-1291. DOI: 10.1007/s13278-012-0079-3

Suh, J. H. (2015). Forecasting the daily outbreak of topic-level political risk from social media using hidden Markov model-based techniques. *Technological Forecasting and Social Change*, *94*, 115-132. DOI: 10.1016/j.techfore.2014.08.014

Tang, Z., Chen, L., Zhou, Z., Warkentin, M., & Gillenson, M. L. (2019). The effects of social media use on control of corruption and moderating role of cultural tightness-looseness. *Government Information Quarterly*. DOI: 10.1016/j.giq.2019.06.001

Terán, L., & Mancera, J. (2019). Dynamic profiles using sentiment analysis and twitter data for voting advice applications. *Government Information Quarterly*. DOI: 10.1016/j.giq.2019.03.003

Theocharis, Y., Barberá, P., Fazekas, Z., Popa, S. A., & Parnet, O. (2016). A bad workman blames his tweets: the consequences of citizens' uncivil Twitter use when interacting with party candidates. Journal of communication, 66(6), 1007-1031.

Times of India (2013) [Online], https://timesofindia.indiatimes.com/assembly-elections-2013/delhi-assembly-elections/AAP-announces-it-will-form-govt-in-Delhi-Kejriwal-meets-LG/articleshow/27782708.cms (Accessed on January 15, 2018).

Times of India (2016) [Online], https://timesofindia.indiatimes.com/city/chandigarh/In-latest-campaign-Congress-pledges-one-job-for-each-household-in-Punjab/articleshow/55696947.cms (Accessed on January 15, 2018).

Times of India (2017) [Online], https://timesofindia.indiatimes.com/elections/assembly-elections/punjab/news/navjot-singh-sidhu-joins-congress-calls-himself-born-congressman/articleshow/56609020.cms (Accessed on January 15, 2018).

Times of India Blog (2016a) [Online], https://blogs.timesofindia.indiatimes.com/toi-editorials/battle-for-punjab-with-akalis-facing-serious-anti-incumbency-next-years-assembly-polls-are-wide-open/ (Accessed on January 15, 2018).

Times of India Blog (2016b) [Online], https://blogs.timesofindia.indiatimes.com/sanjeev-singh-blog/opener-navjot-sidhu-will-give-aap-a-credible-cm-face-in-punjab/ (Accessed on January 15, 2018).

Tribune India (2016) [Online], http://www.tribuneindia.com/news/nation/navjot-sidhu-among-6-nominated-to-rajya-sabha/226112.html (Accessed on January 15, 2018).

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *Icwsm*, *10*(1), 178-185.

Tweetinvi API (2017) [Online], https://www.nuget.org/packages/TweetinviAPI/. (Accessed on January 10, 2018).

Van den Broeck, J., Cunningham, S. A., Eeckels, R., & Herbst, K. (2005). Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS medicine*, *2*(10), e267. DOI: 10.1371/journal.pmed.0020267

Visual Studio 2012 (2012) [Online], https://www.visualstudio.com/vs/older-downloads/. (Accessed January 10, 2018).

Vromen, A., Xenos, M. A., & Loader, B. (2015). Young people, social media and connective action: From organisational maintenance to everyday political talk. Journal of Youth Studies, 18(1), 80-100.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Wu, B., & Shen, H. (2015). Analyzing and predicting news popularity on Twitter. *International Journal of Information Management*, 35(6), 702-711.

Yaqub, U., Chun, S. A., Atluri, V., & Vaidya, J. (2017). Analysis of political discourse on twitter in the context of the 2016 US presidential elections. *Government Information Quarterly*, *34*(4), 613-626.

Yardi, S., & Boyd, D. (2010). Dynamic debates: An analysis of group polarization over time on twitter. Bulletin of science, technology & society, 30(5), 316-327.

Yuan, H., Xu, W., Li, Q., & Lau, R. (2017). Topic sentiment mining for sales performance prediction in e-commerce. *Annals of Operations Research*, 1-24.

Zavattaro, S. M., French, P. E., & Mohanty, S. D. (2015). A sentiment analysis of US local government tweets: The connection between tone and citizen involvement. *Government Information Quarterly*, *32*(3), 333-341.

Zhang, C., Fan, C., Yao, W., Hu, X., & Mostafavi, A. (2019). Social media for intelligent public information and warning in disasters: An interdisciplinary review. *International Journal of Information Management*, 49, 190-207.

Zhang, K., Bhattacharyya, S., & Ram, S. (2016). Large-Scale Network Analysis for Online Social Brand Advertising. MIS Quarterly, 40(4).